# An Efficient Hybrid BFGS-CG Search Direction for Solving Unconstrained Optimization Problems

**Article** *in* Journal of Advanced Research in Dynamical and Control Systems · May 2020

DOI: 10.5373/JARDCS/V12SP2/SP20201161

# An Efficient Hybrid BFGS-CG Search Direction for Solving Unconstrained Optimization Problems

*Kamilu Kamfa[1], Muhammad Yusuf Waziri[2], Ibrahim Mohammed Sulaiman[3], Mohd Asrul Hery Ibrahim[4], Mustafa Mamat[3,*], Siti Sabariah Abas[3]*

[1]*Department of Mathematical Science, Kano University of Science and Technology, Wudil, Nigeria.*

[2]*Department of Mathematical Science, Faculty of Physical Sciences, Bayero, University Kano, Nigeria*

[3]*Faculty of Informatics and Computing Universiti Sultan Zainal Abidin, Terengganu, Malaysia.*

[4]*Faculty of Entrepreneurship and Business, Universiti Malaysia Kelantan, Kelantan, Malaysia*

**Abstract:**Recently, various methods for solving unconstrained optimization problems have been proposed. Most of these methods employ different approach to calculate the search direction $d_k$. Some of the famous search direction includes, Newton method, Quasi Newton method, and Conjugate Gradient method (CG). In thispaper,wedevelopanewhybridmethod which uses CG and BFGS search direction simultaneously under strong Wolfe line search. Various Numerical results have been presented to illustrate the efficiency of the proposed method when comparedwithCGandBFGSmethod.UnderstrongWolfelinesearch,we show that our new algorithm convergesglobally.

**Keywords:**Search direction, FR parameter, Sufficient descent, BFGS, Step size

## Introduction

Consider an unconstrained optimization problem

$$\min_{x \in R^n} f(x) \tag{1}$$

where $f: R^n \to R^n$ is continuously differentiable. There are various methods for solving (1), each method is design using different search direction but with the same iterative formula. Some of the known search direction are Conjugate gradient (CG), Newton method and Quasi-Newton method. The CG method is preferred among the three-search direction due to its simplicity and low memory requirement. This method uses the iterative sequence

$$x_{k+1} = x_k + \alpha_k d_k \tag{2}$$

to generate the sequence of iterate. αk is the step size obtained using either an exact or inexact line search. The exact line search is too expensive and most of the time cannot be used to solve practical problems. An inexact line search was developed by previous researchers like Armijo, Wolfe, and Goldstein [10], to address the cost of using exact line search. Others researchers like Shi [10] have proposed a new line search by extending the Armijo line search, Wolfe also develops a strong Wolfe line search [9]. In summary, majority of the research in this domain preferred the strong Wolfe line search given by

$$f(x_k + \alpha_k d_k) \le f(x_k) + \mu \alpha_k g_k^T d_k \tag{3}$$

$$|g(x_k + \alpha_k d_k)^T d_k| \le -\sigma g_k^T d_k$$

with $0 < \mu < \sigma < 1$. Because it provides a better step size and $d_k$ is the search direction. There are numerous search directions in the literature. The Quasi-Newton method was developed to reduce the cost of using the Newton methods [12, 13]. This search direction uses different approaches such as Davidon-Fletcher-Powell (DFP), Broyden family, Symmetric rank one (SR1) and BFGS formula to approximate the Hessian matrix in the Newton iteration formula see [10,12,14]. The BFGS update is the popular among all the classes which is defined as

$$B_{k+1} = B_k - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} + \frac{y_k y_k^T}{s_k^T y_k} \tag{4}$$

where $s_k = x_{k-1} - x_k$ and $y_k = g_{k-1} - g_k$ In each iteration $B_k$ must satisfy the secant condition

$$B_k s_k = y_k \tag{5}$$



DOI: 10.5373/JARDCS/V12SP2/SP20201161

*Corresponding Author: Mustafa Mamat, E-mail: must@unisza.edu.my*

Article History: Received: Mar 05, 2019, Accepted: May 26, 2020

CG method is another search direction $d_k$ defined as

$$d_k = \begin{cases} -g_k & if \quad k = 0, \\ -g_k + \beta_k d_{k-1} & if \quad k \geq 1, \end{cases} \qquad (6)$$

where $g_k = \nabla f(x)$ is the gradient of f at $x_k$ and $\beta_k$ is a scalar, called conjugate parameter. Different choice of $\beta_k$ leads to different CG methods. Some of the well-known CG method include the Hestenes-Stiefel, Polak-Ribiere-Polyak, The Fletcher-Reeves [5,6,7,8] and the recent version of Rivaie et al. [9], Abashar et al. [3] and Mamat et al. [11]. The CG methods with FR methods converge globally with less computational performance. Although, the PRP and HS methods have goods numerical performance but they do not always converge [6]. Gilbert and Nocedal [7], Touati-Ahmed and storey and Powell have studied PRP method using different line search. Al-Baali's [8] gives a hybrid CG method that combine the PRP and FR method. For recent study on numerical methods, please refer to [15-18].

In this research, we proposed a new hybrid search direction which combines two known search directions. In particular, we combine the CG search direction with a KMAR formula [11] and BFGS method. KMAR formula can be reduced to FR method under certain condition. The anticipation is to improve the performance of FR method while maintaining its global convergent properties. The rest of this paper is organized as follows. In section 2, we present the proposed search direction. In section 3, we studied the convergent analysis of the new method. The numerical results are presented in section 4 follows by the conclusion remark in section 5.

## New Method

Numerous hybrid CG methods have been studied by researches with the aim of improving the CG method as discussed above. Most of these hybrids consist of two or more CG parameter. Basically, they are designed to maintain the good convergence properties of FR method as well as good numerical results of the PRP method. Some researchers including [10] present another alternative approach of defining the hybrid method. This approach combines BFGS and CG search direction to generate the new CG direction. Motivated by his approach, this research proposed another BFGS-CG method using strong Wolfe line search. In what follows, we state our new method. Recently, Kamilu et al. (2015) proposed a new nonlinear CG formula, which has the same numerator as the PRP, HS and RMIL formula. The numerator was retained to give the formula a restart property. The method is defined as follows

$$\beta_k^{KMAR} = \frac{g_k^T(g_k - g_{k-1})}{g_{k-1}^T(g_k + g_{k-1})} \qquad (7)$$

This formula under certain condition can be reduced to FR method. However, its performance is far better than FR method under any line search. Now, combining (4) with BFGS search direction we have

$$d_k = \begin{cases} -B_k g_k + \lambda(-g_k + \beta_k^{KMAR} d_{k-1}) & k \geq 1, \\ -B_k g_k & k = 0. \end{cases} \qquad (8)$$

where $d_{k,} d_{k-1}$ are the current and previous directions and $\beta_k$ is a scalar, called conjugate gradient parameter or rational approximation model (for this research). Below, we present the complete Algorithm of the proposed method.

**Algorithm 1:**KMM4 Algorithm
Initialization. Given a starting point $x_0$, let $B_0 = I_n$for $k = 0$.
Step 1: Terminate if $\|g_k\| \leq 10^{-6}$ or $k \geq 1000$.
Step 2: Compute the search direction using (8).
Step 3: Calculate the step size using strong Wolfe line search (3).
Step 4: Updated $B_k$ using (4) and $x_k$ using(2).
Step 5: Set $k = k + 1$, and go to Step 1.

## Convergence Analysis

In this section we present the global convergence of KMM4 method (the name is based on the initial of the researcher). For an algorithm to convergence it must satisfy the sufficient descent condition and the global convergence properties under strong Wolfe line search. All the proof will be supported with Numerical results generated using different benchmark problem.

**Global Convergence of KMM4 Method**

Based on Algorithm 1 we assume that each search direction dk satisfied the descent condition.

$$g_k^T d_k < 0,\ \forall k \geq 1 \quad (9)$$

such that for any constant $c_1 > 0$,

$$g_k^T d_k < c_1 \|g_k\|^2,\ \forall k \geq 1 \quad (10)$$

Equation (10) is called the sufficient descent condition. The iterate point $x_{k+1}$ satisfies the condition

$$|g_{k+1}^T d_k| < -\sigma g_k^T d_k. \quad (11)$$

Next,weshowthatourproposedmethodsatisfies(10). The following assumption are very important in the convergence analysis of the CG method.

**Assumption A**

1. The objective function $f$ is twice continuouslydifferentiable.
2. The level set $L$is convex andsatisfy

$$c_1\|z\|^2 \leq c_1, z^T F(x) z \leq c_2 \|z\|^2,$$

for all $z \in R^n$and $, x \in L$, where $F(x)$ is the Hessian matrix of $F$.

3. The hessian matrix is Lipschitz continuous at the point $x_*$, that is there exist the positive constant $c_3$ suchthat

$$\|g(x) - g(x_*)\| \leq c_3 \|x - x_*\|$$

for all $x$ in a neighborhood of $x_*$.

**Theorem 3.1:** Let the sequence $\{B_k\}$ be generated by a BFGS formula (4), where $B_k$ is symmetric and positive definite and $y_k^T s_k > 0$ for all $k$. Also, assume that $\{s_k\}$ and $\{y_k\}$ are such that

$$\frac{\|y_k - G_* s_k)\|}{\|s_*\|} \leq \epsilon_k \quad (12)$$

forsomesymmetricandpositivedefinitematrix$G(x_*)$andsequence$\{\epsilon_k\}$with property $\sum_{k=1}^{\infty} \epsilon_k < \infty$. Then

$$\lim_{k\to\infty} \frac{\|B_k - G_* d_k)\|}{\|d_*\|} = 0 \quad (13)$$

and the sequence $\{\|B_k\|\}$, $\{\|B_k^{-1}\|\}$, are bounded.

**Theorem 3.2:** Suppose Assumption A and Theorem 3.1 holds. Then condition (10) is true for all $k \geq 0$.
**Proof**. From (10), we have

$$\begin{aligned} g_k^T d_k &= -g_k^T B_k^{-1} g_k + \eta g_k^T \left( -g_k + \left( \frac{g_k^T (g_k - g_{k-1})}{g_{k-1}^T (g_k + g_{k-1})} \right) d_{k-1} \right) \\ &= -g_k^T B_k^{-1} g_k + \eta \left( -g_k^T g_k + \left( \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} \right) g_k^T d_{k-1} \right) \end{aligned}$$

$$= -g_k^T B_k^{-1} g_k + \eta \left( 1 + \left( \frac{g_k^T d_{k-1}}{\|g_{k-1}\|} \right) \right). \quad (14)$$

The last term can be simplified as follows, from (9) and (11), we have

$$-1 + \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2} \leq \frac{g_k^T d_k}{\|g_k\|^2} \leq -1 - \sigma \frac{g_{k-1}^T d_{k-1}}{\|g_{k-1}\|^2},$$



DOI: 10.5373/JARDCS/V12SP2/SP20201161
**Corresponding Author:* *Mustafa Mamat, E-mail: must@unisza.edu.my*
Article History: Received: Mar 05, 2019, Accepted: May 26, 2020

thus,

$$-\sum_{j=0}^{k}\sigma^j = -1-\sigma\sum_{j=0}^{k=1}\sigma^j \le \frac{g_k^T d_k}{\|g_k\|^2} \le -1+\sigma\sum_{j=0}^{k=1}\sigma^j = -2+\sum_{j=0}^{k}\sigma^j \qquad (15)$$

therefore, using (15) in (14) wehave

$$g_k^T d_k = -g_k^T B_k^{-1} g_k + \eta\left(-2+\sum_{j=0}^{k}\sigma^j \|g_k\|^2\right)$$
$$\le -\lambda_k\|g_k\|^2 + \left(-2\eta + \eta\sum_{j=0}^{k}\sigma^j\right)\|g_k\|^2$$
$$\le c_1\|g_k\|^2$$

where$c_1 = -\left(\lambda_k + \left(2\eta + \eta\sum_{j=0}^{k}\sigma^j\right)\right)$, which is bounded away from zero. Hence, (10) is true. □

**Lemma 3.1.** BasedonAssumptionA,there existpositiveconstants$\omega_1$and$\omega_2$suchthatforany$x_k$and$d_k$with$g_k^T d_k <$ 0,thestepsizeproducedbyalgorithm1willsatisfythestrongWolfelinecondition (3).

**Proof**. Suppose $p_k < 1$, and assuming (3) failed for $p_k \le \frac{p_k}{\tau}$

$$f(x_k - p'd_k) - f(x_k) \le \omega p' g_k^T d_k \qquad (16)$$

Basedonthemeanvaluetheoremwehave,

$$f(x_{k+1}) - f(x_k) = \bar{g}^T(x_{k+1} - x_k)$$

where $\bar{g} = \nabla f(\bar{x})$, for some $\bar{x} \in (x_{k+1}, x_k)$. Now, by using Cauchy-Schwartz inequality, we have

$$\bar{g}^T(x_{k+1} - x_k) = g^T(x_{k+1} - x_k) + (\bar{g} - g_k)^T(x_{k+1} - x_k)$$
$$= g^T(x_{k+1} - x_k) + \|\bar{g} - g_k\|(x_{k+1} - x_k)$$
$$\le g^T(p'd_k) + \mu(p'\|d\|)^2$$

From Assumption (3), we get

$$(\omega - 1)p' g_k^T d_k < p'(\bar{g} - g_k)^T d_k \le M(p'\|d\|)^2$$

this gives

$$p_k \ge \tau p' > \tau(1-\omega)\frac{-g_k^T d_k}{M(p'\|d\|)^2}$$

Substituting the above inequality into (16) will give

$$f(x_k + \alpha_k' d_k) - f(x_k) \le c_2\frac{-g_k^T d_k}{(p'\|d\|)^2} \qquad (17)$$

where $c_2 = \frac{\tau(1-\omega)}{M}$.

**Theorem 3.3**: (Global convergence). Suppose that Assumption A and theorem 3.1 is true.Then,

$$\lim_{k\to\infty} inf\,\|g_2\|^2 = 0 \qquad (18)$$

**Proof**. From the descent property (9) and Lemma 3.1, we get

$$\sum_{k=0}^{\infty}\frac{\|g_k\|^4}{\|d_k\|^2} < \infty \qquad (19)$$

from theorem 3.1, defining $\|d_k\| \leq -c\|g_k\|$. This implies (16) follows. Hence, the proof is completed. □

## Numerical Results

In this section, we present the numerical performance of our proposed method when compared with FR, BFGS, RMIL and AMRI method. All the algorithms are coded in MATLAB R2013b and tested for some well-known benchmark problem with $\|g_k\| \leq 10^{-6}$ as stopping condition. We used the performance profiles introduced by Dolan and More [2] to analyze the numerical results. This profile gives the performance of a solver efficiency and probability of success in a better way. In the following, we give details of the test problems used in this work.

**Table 1:** A list of all the test problem

| Test Function | dimension | Source |
|---|---|---|
| Six-hump camel back | 2 | Andrei [1] |
| Extended Shalow | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Raydan 2 function | 2,4,10,100 | Andrei [1] |
| Extended Rosenbrock | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Extended Himmeblau | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Freudenstein and Roth | 2,4,10,100,500,1000,10000 | Rivaie [9] |
| Extended Beale | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Booth | 2 | Rivaie [9] |
| Treccani | 2 | Rivaie [9] |
| Zettle | 2 | Andrei [1] |
| Extended Penalty | 2,4,10,100,500,1000 | Andrei [1] |
| Extended Trigonometry | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Extended Denschnb | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Hager Function | 2,4,10 | Andrei [1] |
| Fletcher | 2,4,10,100,500 | Andrei [1] |
| Generalized Quartic | 2,4,10,100,500,1000 | Andrei [1] |
| Extended Maratos | 2,4,10,100,500,1000 | Andrei [1] |
| Diagonal | 2,4,10,100,500,1000,10000 | Rivaie [9] |
| Ext White and Holts | 2,4,10,100,500,1000,10000 | Andrei [1] |
| Quadratic QF2 | 2,4,10,100,500 | Andrei [1] |
| Nonscomp | 2,4,10,100 | Andrei [1] |
| Generalized Tridiagonal | 2,4,10,100,500,1000,10000 | Andrei [1] |

Figure 1 and 2 shows that, KMM4 method is effective with a better numerical performance, that is way its curves appear at top and reach 1. FR and BFGS method have good convergent rate but they are numerically poor, hence its curves appears below with 0.82 and 0.77 success respectively. RMIL and AMRI methods fall between the two groups with 0.88 and 0.91 success respectively. The performance of KMM4 method is not surprising because it required two search direction which both of them satisfy the sufficient descent properties independent of any line search.



DOI: 10.5373/JARDCS/V12SP2/SP20201161
*Corresponding Author: Mustafa Mamat, E-mail: must@unisza.edu.my
Article History: Received: Mar 05, 2019, Accepted: May 26, 2020

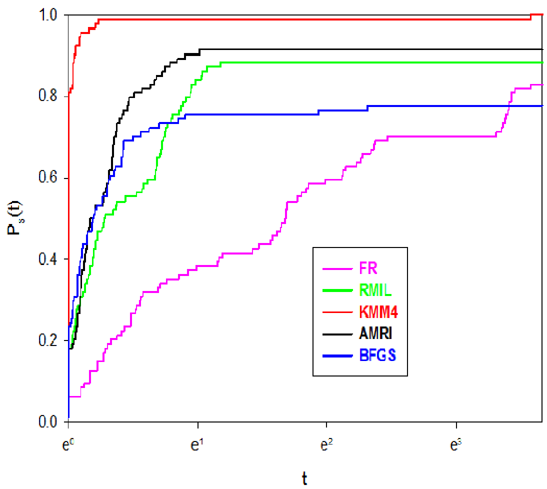


**Figure 1:**PerformanceProfileBasedontheNumberof Iteration Using Strong Wolfe linesearch

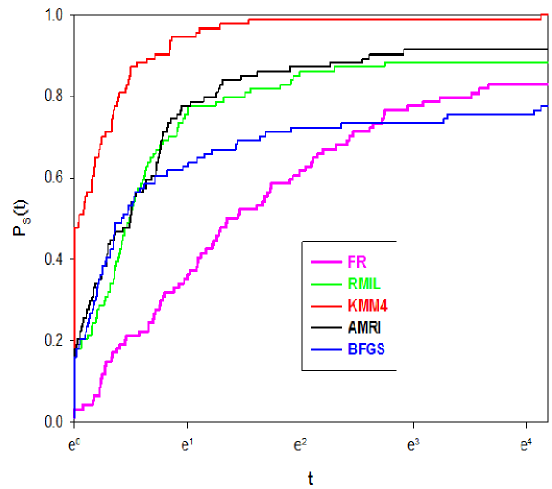


**Figure 2**:Performance Profile Based on CPU Time Using Strong Wolfe line search

## Conclusion

In this paper, we present a new algorithm for solving unconstrained optimization problems. It is a hybrid method that combines the Quasi-Newton search direction with CG method to generate a new search direction under strong Wolfe line search. The idea is to improve the performance of FR method while retaining its good convergent and sufficient descent properties. Numerous standard test function has been used to generate the numerical results. An application of the mathematical approach can be referred to [19].

## Acknowledgment

The author would like to thank the Government of Malaysia for funding this research under the Fundamental Research Grant No. R/FRGS/A0100/01258A/003/2019/00670.

## References

[1] Andrei, N., An unconstrained optimization test functions collection, *Advanced Modelling and optimization,* vol. 10, no. 1,2008, pp. 147-161.
[2] Dolan, E. D. and Moré, J. J., Benchmarking optimization software with performance profiles. *Math. Program.* vol. 91, no. 2, 2002, pp. 201–213.
[3] Abashar, A., Mamat, M., Rivaie, M., and Mohd, I. Global Convergence Properties of a New Class of Conjugate Gradient Method for Unconstrained Optimization. *Applied Mathematical Sciences*, vol. 8, no. 67, 2014, pp. 3307-3319.
[4] Powell, M. J. (1984). Nonconvex minimization calculations and the conjugate gradient method. In *Numerical analysis* (pp. 122-141). Springer, Berlin, Heidelberg.
[5] Rivaie, M., Mustafa, M., Ismail, M., & Fauzi, M. A comparative study of conjugate gradient coefficient for unconstrained optimization. *Aus. J. Bas. Appl. Sci*, vol. 5, 2011, pp. 947-951.
[6] Hager, W. W., & Zhang, H. A new conjugate gradient method with guaranteed descent and an efficient line search. *SIAM Journal on optimization*, vol. 16, no. 1, 2005,pp. 170-192.
[7] Gilbert, J. C., and Nocedal, J. Global convergence properties of conjugate gradient methods for optimization. *SIAM Journal on optimization*, vol. 2, no.1, 1992, pp. 21-42.
[8] Al-Baali, M. Descent property and global convergence of the Fletcher—Reeves method with inexact line search. *IMA Journal of Numerical Analysis*, vol. 5, no. 1, 1985, pp. 121-124.
[9] Rivaie, M., Mamat, M., Leong, W. J., & Ismail, M. A New conjugate gradient coefficient for large scale nonlinear unconstrained optimization. *Int. Journal of Math. Analysis*, vol. 6, no. 23, 2012, pp. 1131-1146.
[10] Ibrahim, M. A. H., Mamat, M., & Leong, W. J. The hybrid BFGS-CG method in solving unconstrained optimization problems. In *Abstract and Applied Analysis,* 2014.
[11] Kamfa, K. U., Mamat, M., Abashar, A., Rivaie, M., Ghazali, P. L. B., & Salleh, Z. Another modified

conjugate gradient coefficient with global convergence properties. *Applied Mathematical Sciences*, vol. 9, no. 37, 2015, pp. 1833-1844.

[12] Waziri, M. Y., Aisha, H. A., & Mamat, M. A structured Broyden's-Like method for solving systems of nonlinear equations. *Applied Mathematical Sciences*, vol. 8, 2014, pp. 7039-7046.

[13] Abdelrahman, A., Mamat, M., Salleh, Z., & Rivaie, M. A modified conjugate gradient method with sufficient condition and conjugacy condition. *Far East Journal of Mathematical Sciences*, vol. 99, no. 4, 2016, pp. 525-540.

[14] Abashar, A., Mamat, M., Rivaie, M., Fauzi, M., & Salleh, Z. A modified DPRP conjugate gradient method for unconstrained optimization. *Far East Journal of Mathematical Sciences (FJMS)*, vol. 97, no. 1, 2015, pp. 31-44.

[15] Sulaiman, I. M., Mamat, M., Puspa, L.G., Hazimi, M. F., Kamilu. K. Benny, P. An Improved Shamanskii Method for Solving Nonlinear Equation. *Jour of Adv Research in Dynamical & Control Systems*, vol. 12, no. 2, 2020, pp. 591-594.

[16] Aini, N. Rivaie, M. Mamat, M. and Sulaiman, I. M. (2019). A Hybrid of Quasi-Newton Method with CG Method for Unconstrained Optimization. *J. Phys.: Conf. Ser*. 1366(012079), 1-10.

[17] Yakubu, U.A., Sulaiman, I.B., Mamat, M., Puspa, L.G., Kamil, K. (2020). The Global Convergence Properties of a Descent Conjugate Gradient Method. Jour of Adv Research in Dynamical & Control Systems, vol. 12, no. 2, 2020, pp. 1011-1016.

[18] Umar, A. O., Sulaiman, I. M., Mamat, M., Hazimi, M .F., Altien, R., Dieby, S. New Hybrid Conjugate Gradient Method for Solving Fuzzy Nonlinear Equations. *Jour of Adv Research in Dynamical & Control Systems*, vol. 12, no. 2, 2020, pp. 585-590.

[19] Ghazali,P.L., Mohd, I, Mamat, M., Ahmad, W.M.A.W. Mathematical Modelling in Family Takaful. Journal of Applied Sciences, 11(19), (2011), 3381-3388.



DOI: 10.5373/JARDCS/V12SP2/SP20201161

**Corresponding Author:* *Mustafa Mamat, E-mail: must@unisza.edu.my*

Article History: Received: Mar 05, 2019, Accepted: May 26, 2020